DOCUMENT RESUME

ED 152 223 IR 005 012

AUTHOR Bierschenk, B.
TITLE A Computer-Based Content Analysis of Interview Texts: Numeric Description and Multivariate Analysis.
INSTITUTION School of Education, Malmo (Sweden). Dept. of Educational and Psychological Research.
PUB DATE Jan 77
NOTE 48p.
JOURNAL CIT Didakometry; n53 Jan 1977

EDRS PRICE MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS Cluster Analysis; *Computer Oriented Programs; *Content Analysis; Discriminant Analysis; Interviews; *Linguistic Patterns; Researchers; Research Methodology; *Research Problems; *Verbal Communication

ABSTRACT

A method is described by which cognitive structures in verbal data can be identified and categorized through numerical analysis and quantitative description. Transcriptions of interviews (in this case, the verbal statements of 40 researchers) are manually coded and subjected to analysis following the AaO (Agent action Object) paradigm. The texts of the interviews are examined first to see whether and to what extent the same words are used in formulating and defining problems, and whether and to what extent the same objects are used. Cluster analysis models are applied, and the ANACONDA (analysis of concepts by means of data-processing) method is employed to initiate an iterative construction of theories and models. A study is made of the grouping of the agents and objects, taking into account how often a particular word occurs. Then both agents and objects are clustered. Logical relations between agents and objects are examined through the AaO paradigm, and objects are transformed into empirically specified concepts by linking objects to respective verbs. Finally, discriminant analysis is used to describe latent patterns of relationships. (WBC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED152223

A COMPUTER-BASED CONTENT ANALYSIS OF INTERVIEW TEXTS: NUMERIC DESCRIPTION AND MULTIVARIATE ANALYSIS

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Bernhard Bierschenk

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

Bierschenk, B. A computer-based content analysis of interview texts: Numeric description and multivariate analysis. Didakometry (Malmö: School of Education), No. 53, 1977.

This report describes the method of approach used in an analysis of the dimensionality of interview texts. By means of cluster analysis models, the interview material has been homogenized. On the basis of these results the relation pattern has then been studied by means of a discriminant analysis. In the final discussion the results are related to (1) the psycholinguistic model and (2) the model of the research process which have guided this research.

Keywords: Psycholinguistics, concept formation, interview data, content analysis, information and documentation, regression analysis.

IR000 5012

CONTENT

	Sid
0. A HANDFUL OF POINTERS TO THE ANACONDA SYSTEM	3
1. CONSTRUCTION OF THEORIES AND MODELS FOR RESEARCH PROCESSES	10
2. NUMERIC DESCRIPTION AND MULTIVARIATE ANALYSIS OF INTERVIEW TEXT	12
2.1 Condensation of registers	13
2.2 Description of manifest relation patterns	16
2.2.1 Clustering of agents	16
2.2.2 Clustering of objects	18
2.2.3 Clustering of blocks	18
2.2.4 A comparison of the results of two cluster analysis methods	19
2.2.5 Clustering of agents & objects	20
2.3 Description of latent patterns of relations	25
3. SUMMARY AND DISCUSSION	32
4. REFERENCES	37
5. APPENDICES	38
1. Mean values and standard deviations for six concept clusters	
2. Probable group affiliation and discriminant values	

0. A HANDFUL OF POINTERS TO THE ANACONDA SYSTEM

This introductory survey is intended to (1) make it easier for the reader to understand our method for an analysis of concept by means of data-processing (ANACONDA) and (2) present some pointers that we hope will give a sufficiently good picture of the individual steps in the cumulative process of developing a theory and a method for studying complex psychological phenomena, such as communication by means of symbols.

We have found when conducting complicated research tasks that the separate reports contribute little to an understanding of the entire problem field. Eac individual piece of work (report) is too often limited by the phenomenon being studied, the empirical or experimental situation involved and data being collected. In addition the writer of the report cannot always communicate to the reader exactly where the individual reports fit into a series of publications, partly due to the difficulty of knowing who the reader is and what prior knowledge he has of the scientific problem with which a specific report is concerned.

The presentation that follows gives the background and framework of the step described in this report.

PROBLEMS

The development of ideas, formulation and solution of problems are behaviours that are closely connected with man's specific ability to express himself verbally. Attempts at exploring and studying not only problem-solving behaviours but also the complex psychological process behind such behaviours are not confined to the field of psychology. For the past twenty years or so such experiments have also been carried out within several other branches of science, such as mathematics, artificial intelligence, information processing and quantitative linguistics. The common factor of the research work done within the various disciplines is the goal of investigating and giving form to invisible symbolic processes and mechanisms. Instead of using simple mathematical formula to try to describe such cognitive operations, researchers working in the fields named above are trying to develop computer programmes to describe them and to test models and theories on what complex psychological structures look like and how processes develop.

CONTENT AND CONTEXT ORIENTED THEORIES

These attempts at developing a theory on the content of the messages by means of which people communicate with each other aim at increasing our understand-

ing of the cognitive structures that are assumed to form the basis for a human being's verbal expressions. By developing our assumptions step by step and continuously testing them, we try to determine their constancy. The question we have asked ourselves is the following:

CAN WE BY MEANS OF NUMERICAL ANALYSIS AND QUANTITATIVE DESCRIPTION IDENTIFY AND CATEGORIZE COGNITIVE STRUCTURES IN VERBAL DATA, SUCH AS INTERVIEW TEXTS?

Considering the scientific debate of recent years on process research and the marked limitations of various assessment schedules as datagathering methods, the successful execution of the task we have set ourselves should make a significant contribution to the research methods that are available to social scientists

GUIDE TO METHOD DEVELOPMENT

Some general facts about the method and model

It is typical for written or spoken text that it is of great complexity and that the type of information to be extracted from the material is seldom or never collected in one single place in the text. If structural relationships are to emerge all the same, the text must be prepared on the basis of certain assumptions. In Bierschenk & Bierschenk (1976) a flow chart is presented stating the individual steps in the development of the method for computer-based content analysis that we suggest. In addition the psycho-linguistic model on which our text analysis is based is also presented.

The different phases are presented below in chronological order.

Basic material

We have restricted our empirical material to apply for the time being to "Information-seeking, documentation and research planning for the R&D work of the school". The design and implementation of an interview study around this theme is described in B. Bierschenk (1974). In this report the assessment scales with which the interviewee were confronted during the interview are al so evaluated.

Impressionistic analysis

One way of evaluating interview texts is to use an impressionistic content analysis. This is based on intuition, insight and impressions, which means that the interpreta tion is based on subjectively found analysis results. Such an analysis is to be found in Annerblom '1974).

Computer-based analysis

A model for searching for information in interview texts, a brief description of preliminary coding rules and some empirical results from the testing of manual allocation of codes on interview text is presented in B. Bierschenk (1976).

Construction of a system of rules

The chances of ready-developed methods being applicable are often linked to the appearance of the material. The attempts at analysis that have been described in the literature and that are of interest to our analysis have been developed with written text as a basis. But since our material is spoken language text (transcribed from recording tape), which when uttered was meant for the ears of the interviewer alone, it became necessary to build up our own system of rules and codes. A preliminary manual and some test results are presented in I. Bierschenk (1974).

Reliability of manual coding

We use the concept "computer-based content analysis" in order to make it plain that we do not intend to develop a method for automatic text analysis. At the same time this means that the basic material must first be coded before mechanical processings of various kinds can be carried out. The success with which two independent coders have been able to apply different coding rules in a similar way is described in Berg (1974).

Representation of manifest language structures

Computer-based content analysis is becoming increasingly used and usable internationally. The demand for programmes and techniques that are adapted to various problem areas is growing. The theory of linguistic representation that we have found most interesting is Schank's "Conceptual Dependency Theory". We came into contact with it after our first coding rules had been worked out and we found that it is in line with our way of treating the text for input into the computer. The way in which the feeding takes place, how identification is specified in the coding and the way in which we build up our lexicon base is described in I. Bierschenk (1975).

Theoretical and psychometric problems

It is very difficult to try to map what is really meant in the research literature by a content analysis. Each content analysis technique is namely based on a specific way of regarding the content in a message. A content analysis presupposes that we can define

that which is to be measured and counted in the analysis. The unit of the analysis in our psycho-linguistic model goes back to the well-known paradigm Agent-action-Object (goal). The components are defined briefly below.

Paradigm of the analysis

Agent Centres of action or goal-seeking entities that make use of resources to reach their goals. This definition also includes groups, organizations or abstractions that fill the function of being an agent.

Action A direct action that is carried out by an agent for the purpose of achieving a goal. The action defines the content of the AaO paradigm.

Object Everything that an action can be directed at or implemented with.

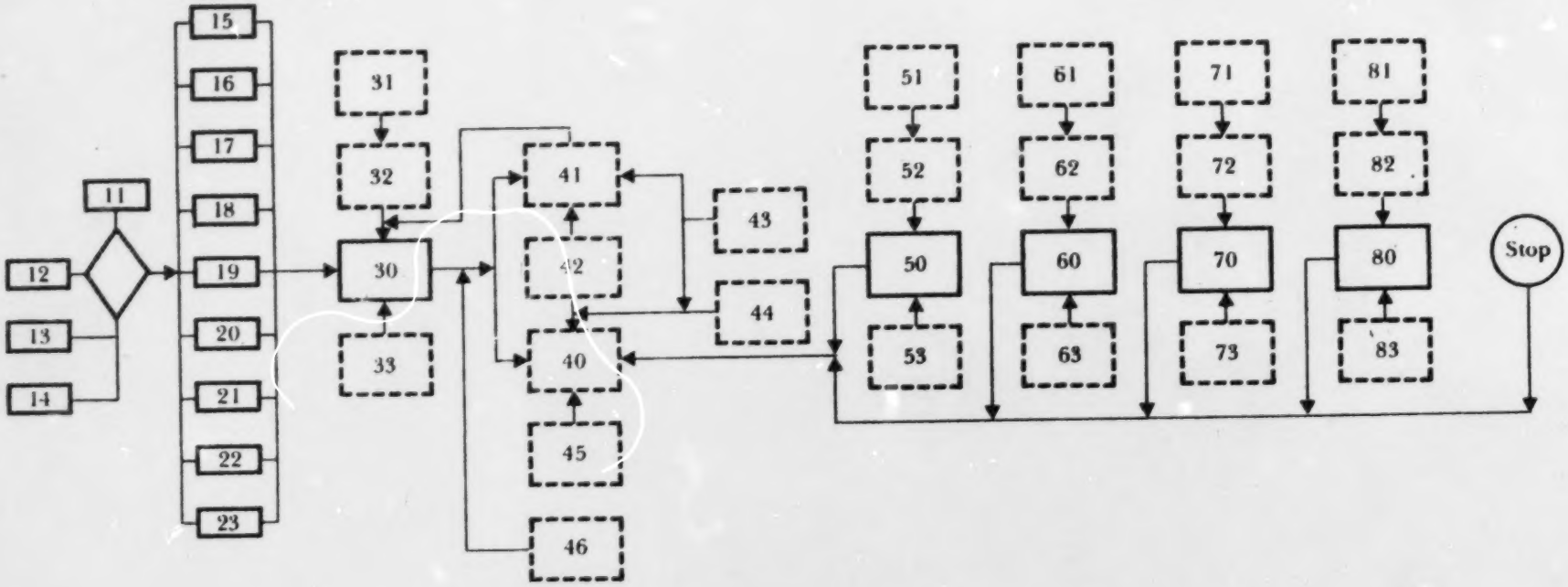
By means of the AaO paradigm, the components that form a natural context, i.e. an observable sentence, are isolated. By this is meant the fundamental form for a statement that is expressed by the $noun_1$-verb-$noun_2$ relationship.

While agent and object (noun) are specified by attribute (adjectival phrases), the verb states the relation between the nouns, i.e. actions, events or state. The order of these basic elements is stated by means of syntax. By using a dictionary and a system of rules we hope to be able to construct theories and models that can be used to describe events.

By means of the linguistic elements to be found in the flow chart in Figure 1, we shall show both how we build up concepts in a given context and the way in which we intend to describe the text numerically and make quantitative analyses. The model in Figure 1 contains three different geometric forms. They have the following meaning:

1. Rectangles which symbolize main concepts which have a code number ending with 0.
2. Rectangles with dotted lines, which symbolize qualifiers or specifiers of various kinds. They have a final number other than 0.
3. A rhomb, which states choice and decision. Very briefly, this means a selection of moods of expression.

The way in which we make use of these possibilities is described in B. Bierschenk (1976).



11	Source of information	21	Interrogation	43	Time	62	Description
12	Negation	22	Supposition	44	Place	63	Classification
13	Tense	23	Volition	45	Modifier of event/state	70	Recipient
14	Mood	30	Agent	46	Circumstance	71	Qualification
15	Condition	31	Qualification	50	Object	72	Description
16	Cause	32	Description	51	Qualification	73	Classification
17	Concession	33	Classification	52	Description	80	Instrument
18	Intention	40	Event/state	53	Classification	81	Qualification
19	Disjunction	41	Copula	60	Direction/localization	82	Description
20	Comparison	42	Clause modifier	61	Qualification	83	Classification

Figure 1. Flow-chart for an analysis of text and the formation of concepts

NUMERIC DESCRIPTION AND QUANTITATIVE ANALYSES

The building up of concepts in a given context presupposes a system of rules that states how different elements are to be linked together and in which order this should take place. We assume, for example, that the relations that exist between nouns and adjevtives and between nouns and verbs reflect such relations as link empirical phenomena with each other. If, for example, we want to state that a set of nouns is modified by a set of adjectives we can formalize this relationship.

Further we assume that nouns get their empirically specified content through adjectives and/or verbs to which they are linked. By scaling adjectives and verbs we can acquire numeric descriptions and quantitative analyses of text. When we have in this way observed similarities or co-variations and defined different properties, we can carry out multivariate analyses in order to determine the position of a particular property in a number of latent dimensions.

Now we will proceed to present individual analyses on the basis of the code structure stated in Figure 1.

Codes 32, 52, 72, 82 and 40

Since in our analysis we take into consideration "syntactic behaviour" and regard both adjective and verb as descriptive concepts, we have decided to scale them. An adjective describes a noun directly, while the verb more indirectly has the same function.

The scaling was carried out by means of seven-point assessment scales, the bi-polar end-points of which are described as pairs of adjectives (1) positive-negative, (2) active-passive and (3) strong-weak. A detailed account of this is to be found in B. Bierschenk (1976) and in Bierschenk & Bierschenk (1976).

REFERENCES TO ANACONDA

Annerblom, M-L. En impressionistisk innehållsanalys av intervjuer med forskare på pedagogiska institutioner i Sverige. /Interviews with researchers in departments of education in Sweden: An impressionistic analysis. /Pedagogisk-psykologiska problem, No. 255, 1974.

Berg, M. Reliabilitetsprövning av en metod för innehållsanalys av intervjutext. /Reliability testing of a method of content analysis to interview texts. / Testkonstruktion och testdata, No. 26, 1974.

Bierschenk, B. Perception, strukturering och precisering av pedagogiska och psykologiska forskningsproblem på pedagogiska institutioner i Sverige. /The perception, structuring and definition of educational and psychological research problems at the departments of education in Sweden. / Pedagogisk-psykologiska problem, No. 254, 1974.

Bierschenk, B. A computer-based content analysis of interview data. Some problems in the construction and application of coding rules. Didakometry, No. 45, 1975.

Bierschenk, B. Teoretiska och psykometriska problem vid en datorbaserad analys av intervjutext. /Theoretical and psychometrical problems in an computer-based analysis of interview texts. /Pedagogisk-psykologiska problem, No. 287, 1976.

Bierschenk, B. & Bierschenk, I. A system for a computer-based content analysis of interview data. (Studia Psychologica et Paedagogica, 32.) Lund: Gleerup, 1976.

Bierschenk, I. Konstruktion av ett regelsystem för en datorbaserad innehållsanalys av intervjutext: Preliminärmanual och några utprövningsresultat. /Construction of rules for a computer-based content analysis of interview texts: A preliminary manual and some evaluation data. /Testkonstruktion och testdata, No. 25, 1974.

Bierschenk, I. Datorbaserad innehållsanalys: Teoretiska och praktiska överväganden. /Computer-based content analysis: Theoretical and practical considerations. /Pedagogisk-psykologiska problem, No. 283, 1975.

1. CONSTRUCTION OF THEORIES AND MODELS FOR RESEARCH PROCESSES

The initial phase of the research process, i.e. the formulation and definition of problems, has been studied empirically. The development of ideas and the formulation of problems are "behaviours" that are closely connected with a human being's specific ability to express himself verbally. As with all other kinds of raw data, the analytical problem in the use of spoken or written text is that the researcher must infer specific events, behaviours or properties that are associated with the "measuring object" of the investigation. Thus the researcher's "messages" concerning the processes of problem perception and problem formulation form the basis of this study.

Treatment of the researchers' verbal statements according to ANACONDA involves an analysis and synthesis of both empirical statements and the relationships between them. This means that we cannot be content with a traditional lexico-graphic listing of words as a base for an approximation of the interview person's implicit models for the research process, but must also take into consideration syntactic order and context. The explicit coding of the basic material that ANACONDA involves means that such material forms a foundation for an iterative construction of theories and models. The flow chart in Figure 1 shows that by starting from elements containing linguistic information, we can build up different analysis units. This presupposes a system of rules that states how different elements are to be linked together (see Bierschenk & Bierschenk, 1976, pp. 90-93). But at the same time it also means that we must be able to handle large amounts of data so that meaningful statistical descriptions and analyses become possible.

The first part of the analysis procedure encompasses steps 1 and 2 (see p. 15) and is carried out in order to study whether and to what extent the researchers make use of the same words when they formulate and define their problems. The frequency distribution shows that we should be able to test an application of cluster analysis models to condense agents that have been used by four or more researchers.

The second part of the analysis procedure encompasses steps 3 and 4 and is carried out in order to study whether and to what extent the same objects are used. The frequency distribution shows that the objects that have been used by four or more researchers can also be condensed by means of some kind of cluster analysis model.

In the third part of the analysis, which encompasses steps 5, 6 and 7, several different cluster analysis models and different amalgamation criteria are

applied when the clusters are formed. But despite relatively good agreement between the results of the different cluster analyses, it seems as if the cluster analyses insufficiently represent the natural groups in the material.

It is in this phase of the analysis process that the research can make use of ANACONDA to initiate an iterative construction of theories and models. According to classical content analysis techniques, we should probably have been content with an explanation of the relations that had emerged and tried to explain them according to the theory of simple associations. The patterns that the result analyses show appear namely to be intuitively meaningful. On the basis of what we "know" about the empirical conditions, however, we cannot allow ourselves to be satisfied with this result.

In the fourth part of the analysis, which encompasses steps 8, 9, 10, 11 and 12, a study is made of the grouping of the agents and the objects when we only take into consideration how often a particular word occurs. The frequency distribution shows that five or more occurrences of the same word is a suitable lower limit.

In the fifth part of the analysis, which encompasses step 13, both the agents and the objects are clustered. These analyses lead to structures that are completely different from those that emerged earlier, partly because more elements are included in the analysis and partly because other criteria have been used in the determination of the elements. At the same time the clusters give the impression of better representing the structure of the agents and objects. But this analysis is again nothing other than a simple expression of similarities (based on the association theory) which are assumed to exist between the units.

Not until the sixth part of the analysis, encompassing step 14, do we exploit the logical relations that exist between the agents and the objects within the framework of the AaO paradigm. By using the verb to give the clause its empirically specified content, we make use of the AaO paradigm in order to reflect the relations that link different empirical phenomena.

In the seventh part of the analysis, encompassing step 15, we make use of the quantitative qualification of the verbs (see Bierschenk & Bierschenk, 1976, pp. 73-89) to give the objects an empirically specified content. By linking the objects to their respective verbs, the objects are functionalized, i.e. they are transformed into empirically specified concepts (see Bierschenk & Bierschenk, 1976, pp. 35-37).

Finally we should perhaps explicitly point out here that we have not made use of all the nuances and variations permitted by the model, but have restricted ourselves to a fundamental structure. It is on the basis of this fundamental structure that we study the dimensionality of the empirically specified concepts by means of a multivariate analysis model.

2. NUMERIC DESCRIPTION AND MULTIVARIATE ANALYSIS OF INTERVIEW TEXTS

The interview method was chosen as the investigatory strategy on the basis of the assumption that the researcher s opportunity to give free and uncommitted answers would provide information of a high validity, at least as long as we can assume that the person interviewed is willing to take part in the investigation. This choice is also motivated by the fact that this method should be more sensitive than a questionnaire with fixed alternative answers, since the interviewees can define their own statements in as differentiated a way as they wish.

In the ANACONDA model different linguistic elements form the building blocks for an empirically specified concept. The dependencies that exist between nouns and adjectives and the relations that exist between nouns and verbs are assumed to reflect the relations that link empirical phenomena.

The first measure taken in building up a concept with an empirical anchorage has been to scale all adjectives and verbs by means of semantic differentials. The present report is a direct continuation of this work. Our purpose is to study the dimensionality of interview texts. In this report we shall describe how we have

1. built up different registers containing linguistic elements
2. used cluster analysis techniques to describe manifest relation patterns based on different registers and
3. used a discriminant analysis technique to describe latent relation patterns.

The analysis results that will be presented are based on the verbal statements of forty researchers, chosen at random from the researcher population (see B. Bierschenk, 1974, pp. 32-44). The entire text material consists of approximately 4 000 pages of text, but the analysis results are based on only 10% of this, i.e. about 400 pages of text. These 400 pages refer to the answers that the researchers gave to four questions (nos. 5, 6, 7 and 8), all of which concern information and documentation problems. These questions have the following wording:

5. In which way have you tried to gain more detailed complementary knowledge?
6. How consistently have you during the formulation process made use of channels of information such as libraries etc?
7. What type of information have you searched for and which search strategy have you used most?
8. Could you say anything about how one should design information searching in order to create ideal conditions for the research process? Have you any suggestions for improvements?

The compilation of lists containing words used by the interview persons when answering questions 5-8 guarantees that our analysis becomes very closely related to the researchers' own language.

2.1 Condensation of registers

Registers can be defined as a formalized arrangement or description of elements (words or speech). If a computer and suitable "Key-Words-In-Context" programme (KWIC) are used, we can build up registers that are very closely related to the text concerned in the analysis. But this is not a sufficient foundation for content analytical processing. We must also be able to use such relations as exist within and between clauses and we must be able to construct categories. A category is usually defined as a number of common attributes. Definition-wise a set of properties is decided upon that are both necessary and sufficient to make it possible to allocate a linguistic element or element complex to a particular category.

If a system of categories is defined on the basis of empirical considerations, it becomes necessary to determine closeness, kinship or similarity (this can also be expressed as distance) for a set of elements.

The interview texts for questions 5-8 have been coded in their entirety in accordance with the AaO paradigm. The scaling of the researchers' actions as registered in verb codes has already been described and in this report the relation patterns between "agents" and between "objects" will be studied. The term agent is used here with the implication of centre of action while the term object stands for means or goal that is the object of an action.

The pre-studies of the appearance of the interview texts have shown that we would get very hollow data matrices. By using different cluster analysis models, we hope to be able to condense both agents and objects, so that the matrices become more complete, i.e. a marking for occurrence in each individual cell.

If cluster analysis techniques are used, it becomes possible to determine agent and object clusters objectively. If the agent clusters are then defined as the measuring objects of the analysis and the concept clusters (determined quantitatively by means of verb links) as the variables of the analysis, statistical analyses become possible. In addition we can draw definable and exact comparable boundaries between all the clusters (categories). This comparison is, however, limited to a single investigation. If a criterion value has been established, the manifest structure within a cluster arrangement becomes completely dependent on the value of the coefficients and is thus no longer exposed to the manipulation of the researcher (see Sokal & Sneath, 1963, pp. 49-53).

Each fusion criterion presupposes certain given mathematical assumptions

and these vary in different cluster analysis models. Since on the one hand we apply several different models and criteria, and on the other have information from analyses carried out with other analysis models we can decide whether the analysis result is meaningful or if it is an artificial structure. Such information is available in Annerblom (1974) and in B. Bierschenk (1974). Cluster analysis techniques can also be used to condense results from cluster analyses.

Before describing the approach used in applying different cluster analysis models for condensing the information contained in the interview texts, we shall give a numeric description of the appearance of the interview texts. The coding of the 400 pages of text that refer to questions 5-8 has resulted in 32 445 units (IBM cards). By defining the AaO paradigm as distinct units, we have been able to build up

1. a noun register that includes agents (A) and objects (O). The register contains 1 634 elements
2. a verb register that contains 1 607 elements and
3. an adjective register that contains 586 elements.

In addition we have constructed

4. a register for noun endings that contains 77 elements
5. a register for verb endings that contains 37 elements and
6. a register for adjective endings that contains 41 elements.

We started a more detailed description of the apperance of the material by studying how the interviewees 1-4 have answered questions 5-8 in the interview. The results of this study are given in Table 1.

Table 1. Numerical description of interview text from interview persons 1 - 4.

Paradigm	Number of cases	%
AaO	864	100.00
aO	86	9.95
Aa	284	34.08
A O	135	16.20
A*	87	10.44
O*	190	22.80

* Registration at first occurrence only.

As can be seen from Table 1, about 60% of the AaO paradigms are incomplete in one way or another. Thus the analysis should begin with a systematic description of each step in the AaO paradigm. If the agents are defined as the mea-

suring objects of the analysis and the objects as the variables of the analysis, we construct a data matrix with 16 530 cells. An empirical control of the number of cells showed that only 342 or 2.27% of the cells contain one or more markings. With this result as a starting point, it became obvious that the material must be condensed and considerable homogenized. For this purpose the following analysis programme was carried out.

Step 1. An analysis of how many agents the 40 interview persons have produced in four questions.

Step 2. An analysis of how the agents are distributed over the interview persons.

Step 3. An analysis of how many objects the 40 interview persons have produced in four questions.

Step 4. An analysis of how the objects are distributed over the interview persons.

Step 5. A cluster analysis of the agents: (1) BMD P01M, in which the agents are treated as variables and (2) BMD P02M, in which the agents are treated as measuring objects.

Step 6. A cluster analysis of the objects: (1) BMD P01M, in which the objects are treated as variables and (2) BMD P02M, in which the agents are treated as measuring objects.

Step 7. A cluster analysis of blocks: (1) BMD P03M, in which agents and objects respectively form one block and the interviewees another and where the threshold value (step lenghts) is placed at .20 and (2) the threshold value at .10 respectively.

Step 8. An analysis of the agents' coincidences with objects, where the coincidences are determined through verbs.

Step 9. An analysis of the number of agents that occur at least five times in the interview material.

Step 10. An analysis of the number of objects that occur at least five times in the interview material.

Step 11. An analysis of the distribution of the number of agents that coincide with such objects as occur at least five times.

Step 12. An analysis of distribution of the number of objects that coincide with such agents as occur at least five times.

Step 13. A cluster analysis of agents and objects respectively: (1) BMD P01M, in which agents and objects respectively are treated as variables.

Step 14. An analysis of the agent clusters' coincidences with the object clusters, where the coincidences are determined by verbs.

Step 15. A discriminant analysis of reduced coincidence matrices.

Since statistical analyses and consequently also cluster analyses assume complete data matrices, the AaO paradigm was studied more closely. Of 3 458 subject-object combinations, only 166 or 4.8% are used by four or more interview persons. Four researchers or 10% was decided upon as a lower limit for a numeric description, since the intention is to be able eventually to carry out

different studies. One way of preparing the interview material so that cluster analyses can be carried out is to standardize the material, so that only the first occurrence of a particular interview person is marked, i.e. $W_{(i,j)} = 1$. All entrances in a matrix become in this way of equal importance. If two matrices are to be compared, the entrances can then easily be rescaled e.g. to the mean 0 variance 1.

An examination of the frequency of use of the agents shows that the analysis material contains 953 agents. Of these, however, only 90 or 9.4% are used by four or more researchers. The examination of the occurrence of the objects shows that the analysis material contains 1 122 different objects. Only 126 or 11.2% are used by four or more interviewees, however. A data matrix of the size 90 x 126 would lead to 11 340 cells, most of which would probably be empty, and a statistical analysis that would be difficult to carry out. For this reason the cluster analyses that will now be described have been made.

2.2 Description of manifest relation patterns

When different cluster models are used to describe relation patterns, the researcher's aim is to homogenize and condense large sets of data. In our case this results in a hierarchical classification. Further classifications can be created by the same technique being used, although with different criteria or also by the use of a different but comparable technique. The clustering results can then be compared. The assessment of whether or not the results obtained are meaningful or not can only be subjective, however.

2.2.1 Clustering of agents

If the interview persons are treated as measuring objects and the agents as variables, we can construct similarity matrices (or distance matrices) and define the distance between two rows in the matrix. The row values are calculated for all the variables. It is assumed that none of the row scores is missing. Most techniques are based on the use of similarity measurements. The values of the similarity coefficients can vary between 1.00 (perfect agreement) and 0 (no agreement at all). One of the cluster analysis programmes that we have used is BMD P01M; the programme is described in Biomedical Computer Programs (Dixon, 1975). In this programme the variables are grouped according to similarity. The measurement used for the association between pairs of variables is "Euclidean distance". This distance is defined as "the square root of the sum of the squared differences between the values for pairs of variables". The summation takes place over all 40 interview persons. In this way a cluster is formed that contains the variables that are most like each other. An amalgamation

algorithm then determines which two clusters are most alike. In this analysis we have made use of "the average linkage algorithm". According to this rule, the mean distance is calculated between a variable in the first cluster and a variable in the second cluster.

The similarity matrices can then be used in several different ways. The cluster analysis technique that is based on Sokal and Sneath's method (BMD P01M) transforms the similarity coefficients to product-moment correlations. The relations that exist between the similarity values and product-moment correlations are given in Table 2.

Table 2. Transformation of similarity values to product-moment correlations.

Similarity values	Product-moment correlations
50.00	.00
55.00	.10
60.00	.20
65.00	.30
70.00	.40
75.00	.50
80.00	.60
85.00	.70
90.00	.80
95.00	.90
100.00	1.00

The correlations shown in Table 2 provide a gauge of the hierarchical classification's adaption to the original similarity matrix (see Anderberg, 1973, pp. 203-204). At the beginning of the analysis each individual agent is regarded as a "cluster". If two clusters fulfil a closeness or distance criterion, they are placed together to form one. The process then proceeds in this way through the whole material. The lowest limit used to merge variables in cluster formation is $r > .30$. The result of the agent clustering is summarized in B. Bierschenk, 1976, Appendix 1:1. With the knowledge we have of the interview texts, the clusters appear to be meaningful from the point of view of interpretation. The fact that relatively many "one-variable clusters" exist indicates heterogeneity in the agents. This can possibly be because we have not in this analysis differentiated between auxiliary verbs (code 41) and verbs (code 40). The construction with an auxiliary verb has a copulative function, i.e. the verb (primarily 'is') binds the agent code with a description or classification, where no action has been described. In this sense it is a question of a qualification. Since this clustering result is only a kind of pre-study, the results of the agent

clustering will not be reproduced in their entirety, but are available at the Department of Educational and Psychological Research in Malmö. In this phase of the analysis process it is also too early to give the clusters general headings.

2.2.2 Clustering of objects

The clustering of objects in which objects are treated as variables, is based on 129 objects. The result of the object clustering is summarized in B. Bierschenk, 1976, Appendix 1:2. In the evaluation of the clustering results the same amalgamation criterion has been used as in the evaluation of the agent clustering. The object clustering shows a considerable reduction in the number of variables, namely from 129 to 48. The cluster structure appears to be meaningful. A sign of the greater homogeneity in the object clusters is that there are several clusters containing three or more variables than is the case in the agent clusters. In addition there are fewer "one-variable clusters". This greater homogeneity can possibly be explained by the selective function of the verbs. It is still too early, however, to give the clusters general headings, so no attempt has been made to find any.

2.2.3 Clustering of blocks

Hartigan's (1972, pp. 123-129) suggestion for a "direct clustering of a data matrix" implies a cluster analysis model that differs in several respects from the BMD P01M and P02M programmes. The analysis model (BMD P03M) is more complex, since both measuring objects and variables are clustered simultaneously. Hartigan (1972, p. 123) writes:

"The principal advantage in this approach is the direct interpretation of the clusters on the data."

Another essential difference between the block clustering technique and the one already described is that the block clustering technique is based on the fact that a comparison of the similarity matrices takes place by means of a calculation of distance, instead of correlations. This distance is represented as a weighed Euclidean distance. The cluster analysis method is based on an attempt to minimize the distance between the matrices, in which the one matrix is the original similarity matrix and the other is a similarity matrix for clusters. Thus what happens is that the differences between the group mean values are tested and the clusters that on a particular level prove to be most like each other with regard to their respective mean vectors or centroids are merged. One distinguishing factor in this method is that the similarity values for the linkages in the most similar clusters can vary (rise and fall) from step to step. Thus it can

happen, for example, that the distance (when the similarity measurement is an expression of distance) between the centroids of certain pairs is less than between another pair that has been combined at an earlier stage. As a result the last linkages take place on a lower level than in the two preceding cases (see Anderberg, 1973, p. 141).

The block cluster analysis was carried out with agents as variables and interviewees as measuring objects. Two different threshold values were used, namely step length .10 and step length .20. But for the sake of completeness all the other step lengths are also given. Together with the two described cluster analyses, this study forms a pilot-study and for this reason we shall not give a more detailed account of the cluster structure but only give a summarized account of the results in B. Bierschenk, 1976, Appendix 1:3. As can be seen in Appendix 1:3, the clustering structures are rather like each other. In addition it emerges clearly which clusters the larger clusters break down into when the limit value is selected more restrictively. In this way the interpretation is also made easier. The most marked difference between the two clustering results given is that the threshold value .20 leads to fewer clusters than is the case when the threshold value .10 is used as a criterion of division. In the first analysis 38 clusters have been formed while the number of clusters in the second analysis is 24. Moreover, in both analyses it appears as if the more abstract agents are clustered and linked together at a relatively late stage in the analysis procedure.

In the next chapter we shall attempt to summarize the results of the agent clustering by comparing the results of the different analysis techniques.

2.2.4 A comparison of the results of two cluster analysis methods

In the introduction it was mentioned that different mathematical assumptions and different criteria of division can have as a result that the use of different cluster analysis techniques leads to different cluster structures. Our purpose in this comparison is to study to what extent there is a kernel of clusters or "natural groups" or at least individual "agents" that remain stable irrespective of which analysis model is used. Thus the purpose is to sort the agents into groups so that the degree of natural associations is high between the agents that have been placed in the same group and low between the members of different groups. In B. Bierschenk, 1976, Appendix 1:4 - 1:5 an account is given for the purpose of comparison of a reorganization of the cluster results from Appendices 1:1, 1:2 and 1:3.

A comparison between the agent clusters resulting from Sokal and Sneath's and Hartigan's methods respectively shows that there are many clusters that

are very like each other. But the analysis result also shows that the "I" reference forms a cluster of its own with comparatively high frequencies. This gave rise to a division of the I-reference in as many distinct elements as there were interviewees. In addition the analysis results suggest that we should not standardize the agents with regard to the interviewees but study the occurence of the agents in the text irrespective of how many interviewees have made use of the same agent. In this way it becomes easier for "natural" agent clusters appearing in the text to be formed.

2.2.5 Clustering of agents & objects

The cluster analysis results presented hitherto are all based on a clustering of agents and objects respectively without taking into consideration the agents´ coincidences with the objects and the reverse. Which agents that coincide with which objects and how often is determined by verb linkages. But our analysis technique is based on the very assumption that the whole AaO paradigm is of importance in an analysis and synthesis of empirical phenomena. In a later stage of the analysis programme (steps 8 - 13) we have studied the occurrence and distribution of the agents that coincide with the objects when the objects are related to the agents via verbs and the reverse. This analysis shows that 888 agents coincide with the objects, while 960 objects coincide with the agents. An analysis of the agents´ and the objects´ frequency distribution shows that a limit can be drawn at frequency 5, i.e. the agents and objects that occur at least 5 times in the entire material (irrespective of the interviewee) should be included in the continued analysis. A new analysis of the distribution of agents and objects with ($N > 5$) shows that there are 222 agents and 192 objects. In order to be able to construct a similarity matrix from a data matrix of the order 222 x 192, it became necessary to expand existing computer programmes. Then a cluster analysis was carried out by means of Sokal and Sneath´s cluster model both for the agents as variables and for the objects as variables. The cluster structures are described in B. Bierschenk, 1976, Appendices 2 and 3. A summary of cluster structure is given there in, Appendix 2:5 - 2:6 and the cluster structure of the objects is summarized in the same publication, Appendix 3:5 - 3:6. As in the earlier analyses, correlations $\geq .30$ has been used as the criterion for amalgamation. A comparison between the clusters described in these Appendices, Boxes 1 and 2, and the clusters that emerge from Boxes 5 and 6 show that the cluster structures have changed completely.

terion values, we can use the cluster analyses to discover quite different structures in our data. The analysis results make it quite plain that if the analysis

procedure is based on the separate parts of the AaO paradigm, we get quite different results than is the case if we take into consideration the relations within the AaO paradigm.

The cluster structure of the agents shows that 227 agents can be condensed considerably, i.e. by more than 50%. This analysis has led to 43 clusters with 2 or more agents and 57 clusters with only 1 agent. A noticeable difference between these agent clusters and those described earlier is that the references to "I" or to a particular project dominate. The greater element of agents also encompasses concepts such as project, institution, library etc. These appear to function as comprehensive terms for persons who can act. Another typical feature in the agent clusters is all the terms for fields of work or functions of various people, e.g. fellow-workers, Ph.D. students, project leaders or archivists. No interpretation will be made now, however.

The cluster structure of the objects shows that 192 objects can be condensed by more than 50%. The object structure contains 53 clusters with 2 or more objects and 31 clusters with only 1 object. The situation is almost the reverse of the agent structure. This indicates that there are greater similarities between the objects. The summary of the agent structure appears to express an action or contact for the purpose of obtaining information (names of libraries and reference organs are to be found, as are reference groups). The summary of the object structure expresses more the actual research work and means of approach, i.e. concepts for the problem area concerned and the abstract words such as problem, programme, report or thesis. From the point of view of interpretation, the clusters appear meaningful, but it is still too early to try to give each individual group a comprehensive heading.

The next step (step 14) in the analysis procedure is an examination of the coincidences of the agent and object clusters. This matrix consists of 64 agents and 53 objects. 64 agent clusters were chosen since, in addition to 43 clusters with 2 or more agents, there are a further 21 clusters with an agent that has a frequency as great as or greater than the lowest frequency in a cluster with 2 or more agents. The resultant coincidence matrix was examined with respect to the number of empty cells. Out of 3 392 cells, there are only 646 or 16% that contain one or more markings. A matrix with such an appearance is hardly suitable for e.g. correlative studies. But since this analysis technique can also be used to condense results from different cluster analyses it is naturally also possible to compress the material further. The result of an iterative assimilation by means of the applied cluster analysis techniques is being worked on at present. But in this account we shall for the moment apply a simpler approach, i.e. all agents and objects with coincidences below 10% are ignored. The result of

this selection was a matrix with 16 agent clusters and 9 object clusters. Since this matrix still contains 34% of the cells with no marking, the matrix has been reduced to a matrix with a size of 14 x 6. By this reduction the number of cells with no observation is reduced to 19%. This matrix provides the necessary condition (that there should not be too many mean value assessments) for us to be able to exploit the dependence that exists within the AaO paradigm in order to give the object an empirically specified meaning. By linking the verb assessments that exist to the objects, each object cluster can get its empirical content. In this way they are transformed to "concept" clusters. If the agent clusters are now cosidered as measuring objects and the concept clusters as variables, we can regard the measuring objects as three independent groups (see Bierschenk & Bierschenk, 1976, pp. 79-89). This type of covariation schedule is suited to a discriminant analysis. But before this analysis is presented and discussed in more detail, we shall present the agent clusters and try to give them a comprehensive denotation. The agent clusters are presented in Box 1 and the object clusters in Box 2.

Box 1. Selected agent clusters

Agents	Description
Cluster 1: Social-psychologically oriented researchers	
ARBASS project	Occupational and social adjustment in mentally retarded adolescents.
Fellow-workers	
STUG project	Studies of generation conflicts
One (33)	Development of independence
I (25)	Individualized teaching in geography
I (35)	Studies of the development of children's personalities in varying residential environments.
Cluster 5: Pupil-oriented researchers	
The FRIS project	Free writing in the middle level of the basic school
I (37)	Preschool - primary school in cooperation
I (17)	Interrupted studies in the basic school
I (28)	The consumer project
Cluster 8: Language-oriented researchers	
Library	
I (36)	Models for bilingual instruction of immigrant children
I (29)	Assessment of essays
Cluster 9: Science-oriented researchers	
I (04)	Problem-solution, mathematics teaching, educational planning

Box 1. cont.

Agents	Description
I (14)	Direction of studies in post-secondary school (physics)
Cluster 11: Cognition-psychology-oriented researchers	
I (01)	Cognitive development (Piaget)
I (18)	Self-instruction methods in teaching the deaf
I (09)	Educational Achievement
I (16)	Goals of adult education organizations - now and in the future
Cluster 12: Researchers interested in methodological problems	
I (27)	Pedagogics in teacher training: problems of content analysis
I (24)	Methodological problems in educational research
I (08)	Statistical methodological problems in educational research
Cluster 13: Researchers interested in programmes for applying influence	
The PUSÄ project	Personality development in backward pupils
The ÖM project	Overall goals
I (05)	Social-psychological aspects of the compulsory in-service training for teachers: programme development
I (32)	Teaching methods in higher education
I (31)	Four-year-olds and their parents (parent education)
I (33)	Social development and social training in the basic school
Cluster 14: Linguistically-oriented researchers	
Bierschenk	Psycho-linguistics
One (29)	Analysis of linguistic structures
I (30)	Teaching methods in German
Cluster 18: Humanistically-oriented researchers	
Researchers	
I (20)	Freedom and equality as basic educational concepts within Western pedagogics
I (06)	Process analysis of non-grading
Cluster 20: Researchers interested in socialization	
I (03)	Studies in the socialization of the school
I (13)	Studies on the internal work of the basic school
Cluster 25: Dissemination of information	
Literature	
Symposium	

Box 1. cont.

Agents	Description	
Cluster 34: Type of information and transference of information		
Book	Reading	Books
Relation	Method	
Norm	Suggestion	
Source	Computer	
Convention	Reference	
Measuring instrument	Content	
Idea	Works	
Problem formulation	PA	
Cluster 44:		
We	researchers, within the project, etc. identification with certain groups	
Cluster 46:		
Person	unspecified	

Three different types of cluster are presented in Box 1, namely such (1) that mainly contain agents referring to project names and persons. . , (2) that exclusively contain agents referring to the dissemination of information, type of information and transference of information and (3) that are relatively unspecified. As can be seen from Box 1, the first type of cluster has been described by means of the problem areas that the researcher in question has mentioned as his point of reference for the interview. The other two types of cluster need no description over and above what can be read from the clusters themselves.

Box 2. Selected objects

Cluster 13: Bibliographical reference
Literature
Reference
Journal
Cluster 14: Research organization
Institute
Project
Seminar
Cluster 29: Discussion of problems
Discussion
Problem

Box 2. cont.

Cluster 46: Channels of information

Library
Symposium
Person
Psychological Abstracts
Reference group
ERIC
Department library
University library
Handbooks
Reviewing organs

Cluster 47: Information on research methods

Design
Summary
Measuring instrument

Cluster 48: Information for demarcation of concepts

Document
Suggestion
Idea

The distinguishing feature of the clusters presented in Box 2 is that content-wise they do not need to be clarified by any description. If these clusters are compared to those discussed in Chapter 2.2.1 - 2.2.4, it emerges clearly that it is only by making use of whole sentences that we can overcome the circumstance that such structures as exist in our texts are broken down in an artificial way.

2.3 Description of latent patterns of relations

Following Chapter 2.2.5 we have arrived at a covariation schedule that is suited to an analysis of more complex relations. By means of a discriminant analysis we shall in this chapter study whether and to what extent the six concept clusters described in Box 2 can be used to separate the groups as far as possible. (For a description of the discriminant analysis model, see Cooley & Lohnes, 1971; Tatsuoka, 1971.)

The use of multivariate techniques usually presupposes many more measuring objects and variables than is the case in this analysis. The more measuring objects that form the foundation for the adaptation of the model to a set of empirical data, the greater the certainty of the model's goodness of fit and subsequently of the generalizability of the results.

Another prerequisite that must be fulfilled is that the measuring object consists of a random sample. This condition cannot be compensated by increasing

the number of measuring objects included in the analysis. This latter condition must be considered fulfilled since the interviewees consist of a random sample. On the other hand 14 measuring objects, 6 variables and 3 groups form a very small set of material which restricts our possibility of generalizing from the results of the analysis.

The decision to use a multivariate analysis model is based on the following considerations:

1. The measuring objects consist not of individual agents, but of groups of agents that are a result of a statistical process of condensation and homogenization.

As has been seen from the analysis procedure described, several hundred observations form the basis for the formation of the clusters that are included in the analysis. This provides a more certain empirical base than would have been the case with individual agents as measuring objects. The same argument applies to the concept cluster included in the analysis.

2. The assessments on which this is based show a high degree of reliability (α max = .86 - .97).

These values indicate reliable measurements. As a result of the small number of measuring objects, however, there remains an element of uncertainty over how well the model fits our data. But this argument is only important for significance testing and generalization. If the discriminant analysis model is used for purely descriptive puposes, the doubts that have been expressed regarding the generalization aspects are of subordinate importance.

Discriminant analyses can be carried out both by using all the variables simultaneously, and by making a step-wise analysis of the relative discrimination ability of each separate variable. Both types of analysis have been made. The computer programmes used are partly Cooley & Lohnes' (1971) MANOVA & DISCRIM, partly the step-wise discriminant analysis programme from SPSS. As far as the simultaneous use of the variables is concerned, both programmes lead to identical results. A more detailed account of the results of the discriminant analysis will be given below against the background of the step-wise, since the comparison between the two types of analysis showed that there is one variable that diminishes the discriminating ability of the discriminant functions.

When three groups exists, two discriminant functions can be formed. The discriminating power of each function is stated by means of the eigen values and canonical correlations that are associated to the respective functions. The eigen values is a gauge of the total variance that exists in the discriminated variables. Table 3 presents the discriminating of the functions.

Table 3. Discriminant functions, eigen value (λ), relative percentage (%) canonical correlations (R)

Function	λ	%	R
1	1.36	75.85	759
2	.43	24.15	549

As can be seen from Table 3, the first function is much more important for the separating than the others. The relationship corresponds to 3:1. The canonical correlation forms another association measurement. R states to what extent a function is related to the "group variables".

The statistical tests that are incorporated into the SPSS programme are Wilk's lambda (Λ) and χ^2. These state the success with which the six concept clusters separate our 3 groups when the variables form a discriminant function. The statistical tests mentioned are given in Table 4.

Table 4. Discriminant functions, Wilk's lambda (Λ), χ^2, df and level of significance (p)

Discriminant function	Λ	χ^2	df	p
0	.30	44.98	10	.00
1	.70	13.28	4	.01

Table 4 shows how both functions are significant and consequently of importance for a discrimination of the groups. Λ stands in a reverse relationship (an inverse measurement) to the power of discrimination. High values on Λ mean that there is no discriminating information of importance left.

The two subsequent discriminant functions are presented together with their respective coefficients (standardized) in Table 5.

Table 5. Discriminant functions (f) and standardized coefficients

Concept-cluster	f_1	f_2
13. Bibliographic reference	.50	.04
29. Discussion of problems	.44	.82
46. Channels of information	.53	-.71
47. Information on research methods	-.17	-.49
48. Information on demarcation of concepts	.32	.33

The coefficients given in Table 5 are weights and can be interpreted in the same way as factor loadings or beta weights in a multiple regression analysis. In this sense the coefficients state which clusters contribute most to differentiation in the respective dimensions. The clusters that are important for the first function are "Bibliographic reference" and "Channels of information", which in addition show high negative weight in the second function.

The ones that are important for the second function are "Discussion of problems" and with a negative sign "Information on research methods". "Information on demarcation of concepts", on the other hand, is equally important for both functions. This result is in agreement with the results presented in B. Bierschenk (1974, p. 64).

The concept cluster "Research organization" has no discriminating power, but rather a reducing effect when the cluster is combined with the others. If we ignore the signs in front of the coefficients, Table 5 shows that clusters 29, 46 and 48 contribute substantially to both functions. The initial values are given in Appendix 1. The two functions described above have been formed in such a way as to make the separation between the groups as great as possible. We shall take a closer look at how far we have succeeded in separating the groups and the extent to which the classification of the individual agent clusters to the respective groups has been satisfactory. Table 6 gives a summary of the result of the classification.

Table 6. Summary of classification results

Group	No. of measuring objects		Group 1	Group 2	Group 3
1	14	n	12	2	0
		%	85.7	14.3	.0
2	14	n	2	9	3
		%	14.3	64.3	21.4
3	14	n	0	2	12
		%	.0	14.3	85.7

Group 1 Evaluation dimension
Group 2 Activity dimension
Group 3 Power dimension

In 78.57% of the cases it has been possible to classify the measuring objects correctly. By classification is meant here the process of determining the probable group a measuring object will belong to when only the measuring object's values in the discriminating variables included in the analysis are

available. But it should perhaps be pointed out that probability assessments of this type presuppose a large number of measuring objects if we are to be able with any certainty to make pronouncements on the probability of an event happening or not. The assessed probabilities on which this classification is based must therefore be considered very uncertain. Bearing this reservation in mind, we can see from Table 6 that the efficiency of the variables in separating the measuring objects is best with regard to groups 1 and 3. The relatively large proportion that is wrongly classified (35.7%) means that as far as the activity dimension is concerned the variables discriminate badly. Table 6 also shows that the wrong classifications in the evaluation and power dimensions are equally large (14.3%).

Another way of studying the classification ability of the discriminant functions is to study the groups to which the agent clusters belong. Since we have a classification of the agent clusters that has been used for derivation of the functions and a comparison between predicted group affiliation, the success of the classification can be measured empirically. The gauge for this measurement is the proportion of correctly classified agent clusters. In Appendix 2:1-2:3 can be found the predicted group affiliation for each individual measuring object and discriminant values. As can be seen from Appendix 2:1, we should, if taking primary consideration to our concept clusters, have predicted group affiliation 2, i.e. the activity dimension for the agent clusters "Linguistically-oriented researchers" and "Dissemination of information".

The classification ability of the discriminant functions for group 3, i.e. the power dimension, is presented in Appendix 2:2. This appendix shows group affiliation 2, i.e. the activity dimension is predicted for the agent clusters "Social-psychology oriented researchers" and "Cognition-psychology oriented researchers". In this "group" there are only 2 incorrect classifications, which means that our six concept clusters function well as prediction variables. Another distinguishing feature in this table (Appendix 2:2) is that the discriminant values for the measuring objects with regard to this dimension are relatively similar.

Finally Appendix 2:3 will show how well the affiliation of the measuring objects to group 2, i.e. the activity dimension, could be predicted. As can be seen from appendix 2:3 it has been more difficult to predict on the basis of the information that is available from the concept clusters the affiliation of the agent clusters to group 2, i.e. the activity dimension. In 35.71% of the cases the agent clusters were incorrectly classified. For the agent clusters "Language-oriented researcher", "Type and transference of information" and "We", affiliation to the power dimension is predicted. Problems of

classification have arisen regarding "Researchers interested in method problems" and "Humanistic-oriented researchers" and whether they belong to group 1 or group 2. This greater uncertainty is also reflected in the greater variation in the discriminant values.

Further information on the differences between groups 1-3 can be obtained from the groups' centroids (stated in Fig. 2 by *) and a graphic presentation of the position of the measuring objects in a two-dimensional discriminant space (group affiliation is stated by the figures 1, 2 or 3).

The centroids in Figure 2 state the mean value of the discriminant values for each group and respective function. As can be seen from Figure 2, Function 1 discriminates well between groups 1 and 3. This function is described primarily by "Bibliographic reference" and "Channels of information". Function 2, the positive pole of which is mainly characterized by "Problem discussion" and the negative pole by "Channels of information" and "Information on research methods", is needed to distinguish as far as possible group 2 from groups 1 and 3. Figure 2 shows that it is easier to differentiate between evaluation and power than between evaluation and activity, or power and activity. But the figure also shows that both functions have a good separating power. This would probably emerge even more clearly if it were not for three so-called outliers. The agent cluster "Dissemination of information" with the discriminant value (1.63, −1.61) in the evaluation dimension and (−1.67, −2.60) in the activity dimension is the one that deviates markedly from the other clusters. The other agent cluster that falls outside its group affiliation is "Methodological problems" with the discriminant values (1.15, 2.00) in the activity dimension.

These circumstances can be studied in more detail by examining the linkages of agent clusters with regard to the values presented in Appendix 1. The agent cluster "Dissemination of information" that includes the agents "Literature" and "Symposium" (see Box 1) will be used to give an example of such an examination. If the agent cluster is related to the concept cluster "Bibliographic reference" (that encompasses the concepts "Literature" "Reference" and "Journal") via the verb linkages, it proves that the actions express a weak positive evaluation (m = 4.59), passivity (m = 3.21) and weak power (m = 3.32).

The second cluster that describes the first function is the concept cluster "Channels of information" (for definition see Box 2). The verbs that relate the agent cluster to this concept cluster express actions that imply a low positive evaluation (m = 4.60) and a certain activity (m = 4.21) though with weak power (m = 3.76).

If this agent cluster's relationship with the concept cluster that defines the positive pole of the second function, we can establish that the actions relating literature and symposium to "Problem discussion" express a negative evaluation (m = 3.46), passivity (m = 2.39) and weak power (m = 3.75).

Against the background of this result, the conclusion we can arrive at is that literature and symposiums do not contribute noticeably to problem discussions. Moreover this agrees entirely with our expectation that the activities connected with symposiums do not express much dynamic, i.e. activity or power. Since this is the case, the actions should not express any very great positive evaluation either.

The negative pole of the second function is defined by the concept cluster "Information on Research Methods" (for definition, see Box 2). The verbs relating this concept cluster to "Literature" and "Symposium" express actions that are somewhat more positive (m = 4.06) in their evaluation. They are also somewhat more active (m = 4.11) and show a marginal increase in power (m = 3.88).

We can draw the following conclusion. The actions that associate dissemination of information via literature and symposium with information on research methods are rather neutral in all three aspects, namely evaluation, activity and power. But it is also plain that information in research methods is to a somewhat greater extent sought and disseminated via literature and symposium than is the case with such information as is of importance for problem discussions.

3. SUMMARY AND DISCUSSION

ANACONDA implies an attempt to make the method for content analysis more objective and more flexible than classical content analyses are. In our case to objectify means that originally subjective functions are transferred to computers, while flexibility means that ANACONDA is supplied with routines that are distinguished by a large capacity for retrieval. The development of an analysis method that is characterized by such qualities demands careful, reliable and valid analyses. This type of method development can only take place step-wise and in interaction with the basic material.

In this report we have described the outcome of the first step in the quantification of the AaO paradigm, in which we have only made use of the empirical qualification of the verb. The continued method development involves a study of each individual element (in Fig. 1) and its relative contribution of information. The most immediate work planned concerns an analysis of the importance of the adjectives and a quantification of the adverbs, plus an analysis of the relative contributions of the adverbs.

One demand that must be made on ANACONDA is that the method must lead to a valid systematization of verbal statements. An attempt to demonstrate the validity of the method by means of the systematization of the researcher's answers to questions concerning information and documentation that exist in the form of the impressionistic content analysis (see Annerblom, 1974) and the evaluation of the assessments on seven-point bipolar assessment scales presented in B. Bierschenk (1974, pp. 63-69) will be described now. It should be possible to compare the compressed summary given with the results of the discriminant analysis and the conclusions presented in that context.

A thorough and systematic check of research publications is the exception rather than the rule. It has also been known for researchers first to gather data and then look for suitable literature. Personal contacts are felt to be the best source of information, but do not appear to play an important part in the researcher's attempts to bring about a supplementary exchange of information. The assessments show that the researchers search primarily for information that will help them to develop an idea, so that the product will be a well-facetted problem, the various facts of which will be suited to a scientific attack. Information for the demarcation of concepts appears to be of a quite special type, since it is not sought together with other information that is of importance for the development of the research strategy. The analysis shows a negative relation between this type on the one hand and on the other, opinions

and interpretations, empirical relations and evidence, norms and conventions, measuring instruments or methods for the processing of data that is to be collected.

The process of problem formulation is highly dependent on the researcher's information-searching behaviour and determination to become acquainted with the research literature in his own field. The impressionistic analysis of the researchers' comments shows that the library is in many cases used because of good personal relations with the library staff (N.B. cluster 46 contains person as an element). Reference organs such as ERIC, Psychological Abstracts (PA) and others form one group that researchers use. It emerges from the comments on the evaluation of reference organs that they "feel dissatisfaction" and that they wish for measures to be taken to improve the quality. ERIC, for example, has attracted little attention (N. B. cluster 46 contains in addition to various types of library the different reference organs mentioned here). Symposiums are attended roughly once a year, but only by certain researchers. Our assessments show that information obtained from reference organs and symposiums is evaluated lowest (with the exception of foreign symposiums).

The methods used in searching for references to literature are unsystematic and employed periodically. Often the researcher starts from references in current literature and searches backwards from these in handbooks, journals and articles. The expectations of obtaining information from libraries are low, however. In the suggestions for improvements a wish is expressed for a better overall view and help in structuring the enormous flow of information. But it is also said that the researchers need not stress their way through masses of literature for fear of missing something. Conversations with other researchers, regular project meetings and seminars are used, on the other hand, for problem discussion and informal literature seminars to stimulate the interest in reading. These problem discussions appear to be the main source of ideas and problem demarcation, since the researchers primarily try to get bibliographic references via different types of channels of information (symposiums, libraries, reference organs, persons and handbooks).

The researchers do not appear to search for information on research methods while the problem discussion is underway or when information for the demarcation of concepts is sought. This result is also supported by all the critical opinions on printed information material. Ideas and suggestions do not seem to be particularly accessible via this type of information. Nor is information on research methods that is available in handbooks and works of reference sought to any great extent. Instead such information is sought mainly from tutors and fellow-researchers.

If the impression given by this account of results is compared with the results given as an example of the outcome of the discriminant analysis together with the comments to be found in connection with Table 5, there can be little doubt about the agreement between the results, namely:

1. Bibliographic references are sought via different types of channels of information.
2. The information the researcher tries to get via problem discussions is different from that which he searches for via channels of information.
3. The researcher seeks information for demarcation of concepts mainly via problem discussions. Information on research methods, on the other hand, is sought neither by the use of different channels of information nor through problem discussions.
4. Information for demarcation of concepts forms a particular type and seems to be negatively related to information on research methods.
5. Dissemination of information in the form of literature and symposiums is related via actions to bibliographical references, channels of information, problem discussion and research methods. The evaluations, the activity and the power these actions express show neutral to negative attitudes.

The problem in connection with an empirical analysis is to choose suitable or strategic parts in a set of data. This cannot take place independently of a relatively explicitly described initial model or theory, however.

The theory on the research process that has guided the collection of the interview material has been described in detail in B. Bierschenk (1974) and the theory of the underlying cognition processes that guided the empirical analysis are discussed at length in Bierschenk & Bierschenk (1976). We shall now illustrate the way in which our empirical results can be introduced into the initial models.

We should perhaps emphasize here that the basic material for this analysis has been limited to the researchers' answers to questions concerning information and documentation. The arguments for this selection have been put forward in several reports but will be repated here. This material has been chosen for the purpose of studying the information-seeking strategies of researchers. But the material was also chosen in the hope that the analysis results would prove to be intuitively meaningful, since the questions on information and documentation are concrete.

An analysis of cognitive processes presupposes that they can be represented, i.e. be made manifest. On the manifest level in our psycholinguistic process

model, we are studying quantitatively linguistic elements and syntax. The results in this report concerning the manifest level consist of numeric descriptions in the form of different observed frequencies. They are reported mainly in Chapter 2.1.

The other level in the model symbolizes relations between concepts. In this sense the AaO paradigm approximates complex cognitive phenomena. The fundamental assumption made here is that knowledge of the direction of the actions is of central importance for a behavioural-(scientific) analysis of the relations between concepts. The analysis result related to this level is described in this report mainly in Chapters 2.2.2 - 2.2.5.

The hypothesis on which the analysis on the next level is based is the following: The central importance of directed activity emerges from the verb's function in the determination of the nature of the AaO paradigm. Nouns (agents, objects) in the clause form distinct units ('mneme) that are operationalized by means of verbs that lose their meaning, i.e. they are stored as abstract relations between nouns. Based on this hypothesis selected objects are in Chapter 2.3 transformed to empirically specified concepts.

On the last level in our psycholinguistic process model the assumption is made that the concepts (based on observable clauses) that have been created consist of empirical evidence, on which plans are developed and functionalized, i.e. become strategies. The result analyses concerning this level are described in Chapter 2.3, primarily Table 5.

The result of the cognition-psychological analysis will now be utilized to make explicit how researchers perceive problems and by which methods they try to solve problems, i.e. achieve scientific goals. It must be possible to formalize each problem, i.e. we must be able to formulate hypotheses. It must be possible to give methods a concrete form and instrumentalize goals, i.e. we must be able to develop techniques by means of which scientific goals can be achieved. ('A more detailed model and description may be found in B. Bierschenk, 1974, pp. 13-25).

The results presented in points 1 - 5 above will now be introduced into this model.

The researcher's plan for solving his information problem contains intentions and goal notions, plus an idea of which means can be used to achieve goals, i.e. means-goal hierarchies. The intention is to get in principle two types of information '1) for demarcation of concepts and '2) about research methods.

The strategy 'means) that has been designed for the first type is problem discussion (discussion seminars, project meetings, informal literature semi-

nars). The strategy designed for the second type is to a certain extent bibliographic information-seeking and visits to international symposiums. But primarily tutors and fellow-researchers are asked. Different strategies are used for getting information about demarcation of concepts and information on research methods respectively. Since both types of information are negatively related to each other, we can draw the conclusion that the information-seeking strategy used is related to the type of information sought.

Instrumentalization, i.e. the technical systems available for channeling information, is used to a certain extent to obtain bibliographic references, i.e information about information. But the actions that form the building stones of the researchers´ information-seeking strategies express a neutral to negative attitude.

4. REFERENCES

Anderberg, M.R. Cluster analysis for applications. New York: Academic Press, 1973.

Annerblom, M-L. En impressionistisk innehållsanalys av intervjuer med forskare på pedagogiska institutioner i Sverige. / Interviews with researchers in departments of education in Sweden: An impressionistic analysis. / Pedagogisk-psykologiska problem, No. 255, 1974.

Bierschenk, B. Perception, strukturering och precisering av pedagogiska och psykologiska forskningsproblem på pedagogiska institutioner i Sverige. / The perception, structuring and definition of educational and psychological research problems at the departments of education in Sweden. / Pedagogisk-psykologiska problem, No. 254, 1974.

Bierschenk, B. En datorbaserad innehållsanalys av intervjutext: Numerisk beskrivning och multivariat analys. / A computer-based content analysis of interview texts: Numeric description and multivariate analysis. / Pedagogisk-psykologiska problem, No. 307, 1976.

Bierschenk, B. & Bierschenk, I. A system for a computer-based content analysis of interview data. (Studia Psychologica et Paedagogica, 32) Lund: Gleerup, 1976.

Bierschenk, I. Konstruktion av ett regelsystem för en datorbaserad innehållsanalys av intervjutext: Preliminärmanual och några utprövningsresultat. /Construction of rules for a computer-based content analysis of interview texts: A preliminary manual and some evaluation data. / Testkonstruktion och testdata, No. 25, 1974.

Bierschenk, I. Datorbaserad innehållsanalys: Teoretiska och praktiska överväganden. / Computer-based content analysis: Theoretical and practical considerations. / Pedagogisk-psykologiska problem, No. 283, 1975.

Cooley, W.W. & Lohnes, P.R. Multivariate data analysis. New York: Wiley, 1971.

Dixon, W.J. (Ed.) Biomedical computer programs. Berkeley: University of California Press, 1975.

Hartigan, J.A. Direct clustering of a data matrix. Journal of the American Statistical Association, 1972, 67 (337), 123-129.

Nie, N.H., Hull, C.H., Jenkins, J.G., Steinbrenner, K. & Bent, D.H. Statistical package for the social sciences (2nd ed.) New York: McGraw-Hill, 1975.

Sokal, R. & Sneath, P.H. Principles of numerical taxonomy. San Francisco: Freeman, 1963.

Tatsuoka, M.M. Multivariate analysis: Techniques for educational and psychological research. New York: Wiley, 1971.

5. APPENDICES

1. Mean values and standard deviations for six concept clusters.
2. Probable group affiliation and discriminant values.

Table 7. Mean values and standard deviations for 14 agent clusters, 6 concept clusters and 3 assessment dimensions

	Bibliographical reference			Research organization			Discussion of problems			Channels of information			Information on research methods			Information for demarcation of concepts		
	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
1	4.12	3.65	3.77	4.65	5.34	4.64	4.31	3.69	3.78	4.34	4.78	4.21	3.09	2.95	3.76	4.75	3.23	3.77
2	4.20	4.54	3.87	4.17	4.00	4.05	4.60	4.00	3.86	4.23	4.55	3.92	4.50	5.15	4.15	4.38	4.37	4.19
3	4.29	4.21	3.78	4.17	4.00	4.05	4.55	3.50	3.76	4.18	4.06	4.00	3.35	2.37	3.73	4.03	3.17	3.52
4	4.18	4.60	4.01	4.17	4.00	4.05	4.77	3.73	3.95	4.11	3.89	3.84	4.06	4.11	3.86	4.28	4.19	4.04
5	4.18	4.63	4.23	4.17	4.00	4.05	4.06	2.59	3.37	4.05	4.09	4.02	3.56	3.66	3.80	5.35	5.05	5.04
6	4.38	4.19	3.84	4.43	4.53	4.01	4.23	4.84	3.85	3.97	4.01	3.95	4.10	4.16	3.83	4.39	5.28	4.22
7	4.24	4.37	4.01	3.85	3.23	3.78	4.31	3.69	3.78	4.07	4.27	3.92	4.43	4.63	3.98	4.52	5.23	4.11
8	4.65	5.63	4.27	4.17	4.00	4.05	4.31	3.69	3.78	4.15	3.88	3.55	4.95	5.71	4.60	4.20	4.24	3.77
9	4.35	4.46	3.91	4.37	4.45	4.06	4.30	4.27	4.09	4.11	4.39	3.94	4.35	4.27	3.97	4.32	3.98	3.77
10	4.15	4.15	3.86	3.91	3.59	4.36	4.31	3.69	3.78	4.23	4.51	3.82	4.06	4.11	3.88	3.93	4.57	3.99
11	4.59	3.21	3.32	4.03	3.21	4.07	3.46	2.39	3.75	4.60	4.21	3.76	4.06	4.11	3.88	5.22	4.57	3.82
12	4.50	2.46	2.77	4.17	4.00	4.05	4.86	4.71	4.14	4.17	4.20	3.86	4.23	4.84	3.85	5.22	4.57	3.82
13	4.81	3.93	3.99	4.59	5.22	4.36	4.06	2.59	3.73	4.22	3.93	3.73	4.07	3.25	3.42	4.57	4.23	3.95
14	4.78	4.66	4.07	4.23	2.39	3.14	4.23	4.24	3.53	4.02	4.04	3.69	4.03	4.19	3.69	4.06	2.59	3.37
$\bar{m}$	4.38	4.19	3.84	4.17	4.00	4.05	4.31	3.69	3.78	4.18	4.20	3.86	4.06	4.11	3.89	4.51	4.23	3.96
s	.25	.77	.40	.39	1.05	.45	.41	.89	.26	.16	.28	.19	.54	.99	.30	.48	.82	.41

Agent clusters

1 Social-psychologically oriented researchers
2 Pupil-oriented researchers
3 Language-oriented researchers
4 Science-oriented researchers
5 Cognition-psychology-oriented researchers
6 Researchers interested in methodological problems
7 Researchers interested in programmes for applying influence
8 Linguistically-oriented researchers
9 Humanistically-oriented researchers
10 Researchers interested in sozialization
11 Dissemination of information
12 Type of information and transference of information
13 We
14 Person

Dimensions

1 positive/negative
2 active/passive
3 strong/weak

Table 8. Evaluative dimension with poles positive-negative

Agent cluster	Actual group	Highest probability group	P(G |X)	2nd highest group	P(G |X)	Discriminant scores f_1	f_2
1	1	1	.97	2	.03	1.40	.91
2	1	1	.77	2	.21	.84	.25
3	1	1	.93	2	.05	.95	1.11
4	1	1	.91	2	.06	.79	1.16
5	1	1	.93	2	.05	.82	1.28
6	1	1	.66	2	.19	.31	.79
7	1	1	.62	2	.28	.41	.41
8	1*	2	.52	1	.45	.62	- .40
9	1	1	.63	2	.31	.50	.25
10	1	1	.51	2	.43	.47	- .05
11	1*	2	.57	1	.43	1.63	-1.61
12	1	1	.99	2	.01	1.75	1.48
13	1	1	.80	2	.20	1.15	- .04
14	1	1	.76	2	.19	.62	.55

For description of agent clusters, cf Table 7.

P(G |X) : Probability that a member of the predicted group really is a member of that group provided that the actual group membership is known.

Table 9. Power dimension with poles strong-weak

Agent cluster	Actual group	Highest probability group	P(G | X)	2nd highest group	P(G | X)	Discriminant scores f_1	f_2
1	3*	2	.65	3	.28	- .36	- .64
2	3	3	.73	2	.22	- .69	.19
3	3	3	.75	2	.24	- .92	- .20
4	3	3	.81	2	.12	- .66	.69
5	3*	2	.50	1	.27	.01	- .03
6	3	3	.70	2	.22	- .55	.38
7	3	3	.69	2	.25	- .62	.18
8	3	3	.97	2	.03	-1.52	.52
9	3	3	.73	2	.20	- .60	.40
10	3	3	.90	2	.09	-1.00	.46
11	3	3	.98	2	.02	-1.74	.46
12	3	3	.99	2	.01	-1.72	.75
13	3	3	.95	2	.05	-1.27	.49
14	3	3	.99	2	.01	-1.71	.98

For description of agent clusters, cf Table 7.

P(G | X) : Probability that a member of the predicted group really is a member of that group provided that the actual group membership is known.

Table 10. Activity dimension with poles active-passive

Agent cluster	Actual group	Highest probability group	P(G | X)	2nd highest group	P(G | X)	Discriminant scores f_1	f_2
1	2	2	.93	1		.60	-1.98
2	2	2	.87	1		1.13	-2.06
3	2*	3	.55	2		- .38	.24
4	2	2	.40	3		- .18	.16
5	2	2	.86	3		- .05	-1.23
6	2*	1	.99	2		1.15	2.00
7	2	2	.64	1		.76	- .84
8	2	2	.84	1		.25	-1.13
9	2*	1	.52	2		1.00	- .69
10	2	2	.80	1		.86	-1.44
11	2	2	.67	3		-1.67	-2.60
12	2*	3	.76	2		- .74	.20
13	2*	3	.74	2		-1.34	- .93
14	2	2	.60	3		- .28	- .43

For description of agent clusters, cf Table 7.

P(G | X) : Probability that a member of the predicted group really is a member of that group provided that the actual group membership is known.